PROCEEDINGS OF TDWG

Biodiversity Information Science and Standards 1: e20518
doi: 10.3897/tdwgproceedings.1.20518

OPEN ACCESS

Conference Abstract

Use of Online Species Occurrence Databases in Published Research since 2010

Joan E Ball-Damerow‡, Laura Brenskelle§, Narayani Barve§, Pam Soltis§, Raphael LaFrance§, Arturo H. Ariño|, Robert Guralnick§

‡ Field Museum of Natural History, Chicago, United States of America
§ University of Florida, Gainesville, United States of America
| University of Navarra, Navarra, Spain

Corresponding author: Joan E Ball-Damerow (joandamerow@gmail.com)

Received: 22 Aug 2017 | Published: 22 Aug 2017

Citation: Ball-Damerow J, Brenskelle L, Barve N, Soltis P, LaFrance R, Ariño A, Guralnick R (2017) Use of Online Species Occurrence Databases in Published Research since 2010. Proceedings of TDWG 1: e20518. https://doi.org/10.3897/tdwgproceedings.1.20518

Abstract

Museums and funding agencies have invested considerable resources in recent years to digitize information from natural history specimens and contribute to online species occurrence databases. Such efforts are necessary to reap the full benefits of irreplaceable historical data by making them openly accessible and allowing the integration of collections data with other datasets. However, recent estimates suggest that still only 10% of biocollections are available in digital form. The biocollections community must therefore continue to justify and promote digitization efforts, particularly for high-diversity groups with large numbers of specimens, such as invertebrates. Our overarching goal is to determine how uses of biodiversity databases have developed in recent years, as more data has come online. To this end, we present a bibliometric analysis of published research to characterize uses of online species occurrence databases since 2010.

Relevant papers for this analysis include those that use online and openly accessible primary occurrence records, or those that add data to an online database. Google Scholar (GS) provides full-text indexing, which was important to identify data sources that often appear buried in the methods section of a paper. Our search was therefore restricted to GS. We drew a list of relevant search terms and downloaded all records returned by each

© Ball-Damerow J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

search (or the first 500 if there were more) into a Zotero reference management database. About one third of the 2500 papers in the final dataset were relevant. Three of the authors with specialized knowledge of the field characterized relevant papers using a standardized tagging protocol based on a series of key topics of interest. We developed a list of potential tags and descriptions for each topic, including: database(s) used, database accessibility, scale of study, region of study, taxa addressed, general use of data, other data types linked to species occurrence data, data quality issues addressed, authors, institutions, and funding sources. Each tagged paper was thoroughly checked by a second tagger.

The final dataset of tagged papers allow us to quantify general areas of research made possible by the expansion of online species occurrence databases, and trends over time. For example, preliminary results on a subset of the papers indicate that the most common uses of online species occurrence databases have been: (a) to determine trends in species richness or distribution; (b) to describe a new database; and (c) to assist in developing species checklists or taxonomic studies. Studies addressing plants have generally been more prevalent than those concerning both vertebrates and invertebrates. However, while the number of plant and vertebrate studies have remained relatively constant in recent years, invertebrate studies are increasing. We also address the importance of both proper citation of databases and use of approaches to improve data quality issues involving errors and biases. The most common aspects of data quality addressed were to check for currently valid names, spatial errors, and to exclude certain unsuitable records. Finally, we identify more integrative studies that incorporate multiple data types, and determine whether these uses are enabled by collaborations.

Overall, our presentation demonstrates initial trend results for over 100 specific tags associated with 13 topics of interest, and network analyses of authors and institutions for relevant papers. We also outline the downstream utility of our dense tagging approach for understanding domain-wide trends, and the potential for developing machine-learning approaches to more efficiently characterize certain aspects of published research.

Keywords

biodiversity data, species occurrence database, online database, natural history specimens, observation records, data quality

Presenting author

Joan E. Ball-Damerow